Data Augmentation techniques in time series domain: a survey and taxonomy

Abstract

With the latest advances in deep learning-based generative models, it has not taken long to take advantage of their remarkable performance in the area of time series. Deep neural networks used to work with time series heavily depend on the size and consistency of the datasets used in training. These features are not usually abundant in the real world, where they are usually limited and often have constraints that must be guaranteed. Therefore, an effective way to increase the amount of data is by using data augmentation techniques, either by adding noise or permutations and by generating new synthetic data. This work systematically reviews the current state of the art in the area to provide an overview of all available algorithms and proposes a taxonomy of the most relevant research. The efficiency of the different variants will be evaluated as a central part of the process, as well as the different metrics to evaluate the performance and the main problems concerning each model will be analysed. The ultimate aim of this study is to provide a summary of the evolution and performance of areas that produce better results to guide future researchers in this field.